UPS-NOI/POIR-T-1

BEFORE THE
POSTAL RATE COMMISSION

POSTAL RATE AND FEE CHANGES, 2000	DOCKET NO. R2000-1

RECEIVED
AUG 21 10 27 AM '00
POSTAL RATE COMMISSION
OFFICE OF THE SECRETARY

TESTIMONY OF UNITED PARCEL SERVICE
WITNESS KEVIN NEELS IN RESPONSE TO
NOTICE OF INQUIRY NO. 4 AND PRESIDING
OFFICER'S INFORMATION REQUEST NO. 19

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH

My name is Kevin Neels. I have previously submitted testimony in this proceeding on the volume variability of mail processing labor costs (UPS-T-1) and on purchased transportation costing (UPS-T-3). My biography is set forth in that testimony. See Tr. 27/12773-74. I have also submitted rebuttal testimony on the volume variability of purchased transportation costs (UPS-RT-1).

SCOPE AND PURPOSE OF MY TESTIMONY

On August 2, 2000, the Commission issued two requests for additional information relating to mail processing costs and the study of mail processing cost variability submitted by Postal Service witness Bozzo:

1. Notice of Inquiry No. 4 ("the Notice") invited interested parties to submit statistical information and analyses comparing the model specification presented by Dr. Bozzo to other alternative specifications.

2. Referring to my calculations of error rates in the MODS data used by Dr. Bozzo (presented in my direct testimony, UPS-T-1, and in my responses to interrogatories), Presiding Officer's Information Request No. 19 asked me to extend my calculations to encompass types of errors I had not previously considered, and to comment on the extent to which the processes giving rise to these errors may have infected apparently error-free observations.

This testimony constitutes my response to these requests.

NOTICE OF INQUIRY NO. 4

1. Summary

As the Notice states (at page 1), in Docket No. R97-1 Dr. Bradley reported the results of a number of statistical tests comparing his preferred models of mail processing cost variability to a number of alternative specifications. In response to Notice of Inquiry No. 4 in that proceeding, these tests were extended to a broader range of alternative specifications. Noting that the record in the current proceeding does not address the same range of model specifications that were evaluated in R97-1, the Notice invites interested parties to test Dr. Bozzo's models against a number of alternative specifications, and to comment on the appropriate basis for selecting a preferred model from the set of available alternatives. Notice at 1-2. It asks interested parties to conduct and present the results of formal statistical tests, but also asks whether the results of such tests should constitute the sole basis for selecting a preferred specification. Id. at 3.

The Notice identified two alternative versions of Dr. Bozzo's model. The first, which was identified as "Model A," contained a complete set of facility-specific fixed effects. This was the specification preferred by Dr. Bozzo. The second model, which was identified as "Model B," lacked facility-specific fixed effects but contained a complete set of time period-specific fixed effects. The Notice requested for each model a test of the null hypotheses (a) that the fixed effects were equal (and hence could be replaced with a single constant term) and (b) that they were not fixed, but rather independently and identically distributed random variables.

In addition to these formal statistical tests, the Notice invited discussion of a number of related topics. It asked (1) whether the requested test results provided a sufficient basis for the selection of Model A over alternatives such as the "pooled" or "random effects" models; (2) whether analogous test results for Model B provided a sufficient basis for its selection; (3) whether Models A and B were nested within one another, and whether the statistical tests requested provided grounds for the selection of one model over another; and (4) whether, apart from the statistical results, there may be theoretical reasons for selecting one model over another.

2. Hypothesis Testing Framework

The models described in the Notice as A and B are not nested, in the sense that neither is a special case of the other. For this reason, there is no direct statistical test leading to the selection of one and the rejection of the other. It is possible, however, to specify a more general model that includes both Model A and Model B as special cases. In the Commission's notation, such a general model would take the following form:

$$y_{it} = \alpha_i + \gamma_t + x_{it}\beta + \varepsilon_{it} \qquad (1)$$

where the α_i represent facility-specific fixed effects and the γ_t represent time period-specific fixed effects. I will refer to this general model as "Model C."

Tests involving Model C can shed some light on the choice between Models A and B. If it were the case, for instance, that one could reject Model A in favor of Model C but could not reject Model B in favor of Model C, this would suggest that Model B would be the better specification. In effect, in such a situation the testing sequence would start with the general model and lead eventually to the more parsimonious specification provided by Model B. However, if results compel the rejection of both A and B in favor of C, the clear implication would be that C was the better model and should be chosen over either of the alternatives.

The Notice points out that it might not be possible to include a complete set of time period specific effects in Dr. Bozzo's model because of collinearity with the variables it contains. Dr. Bozzo's model contains five variables that vary only across time and not across facilities: three seasonal dummy variables, a time trend, and a time trend squared. Perfect collinearity can be avoided by omitting these variables from the specification. In the regression runs reported below, I have eliminated them.[1] This means that the OLS model against which Model A is tested differs from the OLS model against which Model B is tested.

Statistical tests of the type requested in the Notice assume that one of the two models under consideration is correctly specified. If these models are misspecified – in particular, if both omit significant independent variables – coefficient estimates for both of the candidate models will be biased, and tests distinguishing between them will be unreliable. In the present circumstances, this precondition places important limitations on the value of the tests that have been requested. I will discuss this point and its implications in more detail below.

1. Dr. Bozzo's data set contains twenty-four time periods, only nineteen of which appear in his regression sample. The first of the other five time periods is dropped from the analysis because it coincided with significant restructuring of Postal Service systems. The other four are used to calculate the lagged values he requires. Thus, adding a full set of time period-specific effects to Dr. Bozzo's model would require nineteen terms if no constant term were present, and eighteen if a constant term were present. Restrictions necessitated by collinear variables require the elimination of an additional five terms.

3. Statistical Methodology

In estimating these models, I have followed the procedures described by Dr. Bozzo in USPS-T-15 and used a feasible generalized least squares (FGLS) procedure that corrects for first order serial correlation. In the course of responding to the Notice, I uncovered a number of errors in Dr. Bozzo's original methodology. Because the programs I had used in my earlier testimony were designed to replicate Dr. Bozzo's results, they incorporated some of the same errors. I was able to correct some of the errors in Dr. Bozzo's work, but time constraints have prevented me from reworking all of his analysis. In the results presented below, I note the instances in which there remain uncorrected errors.

Dr. Bozzo's FGLS procedure consists of three steps. First, he estimates the coefficients of the model ignoring the possibility of serial correlation. Second, from the residual vector produced in this way he computes an autocorrelation coefficient. In the third and final step, he transforms the data to eliminate the serial correlation. This transformation involves multiplication of all variables for the first observation in each run of data by $\sqrt{1-\rho^2}$, where ρ is the estimated autocorrelation coefficient.[2] He transforms subsequent observations by subtracting from each variable ρ times its value in the previous time period. Researchers using this procedure often simply drop the first observation from their analysis samples. Dr. Bozzo describes this as the "textbook

2. A "run" of data is a set of contiguous non-missing observations for a specific facility. If useable data are present for a particular time period for a specific facility, that would represent a single run. A gap in the middle would divide the data into two runs. A second gap could divide the data into three runs. Dr. Bozzo applies the $\sqrt{1-\rho^2}$ transformation to the first observation in each run.

alternative," but states that he chose not to follow it because doing so would have "adverse consequences for the statistical efficiency of the estimates."[3] He uses this general procedure to estimate his pooled, random effects, and fixed effects models.

In estimating his fixed effects model, Dr. Bozzo uses a preprogrammed function in TSP. Rather than including explicitly in the model a dummy variable for each facility, this procedure uses a computational shortcut in which each variable in his model is expressed in terms of deviations from its facility means.[4] Dr. Bozzo first runs this fixed effects estimator ignoring autocorrelation. He then computes the autocorrelation coefficient, applies the ρ transformation described above, and reruns the fixed effects estimator on the transformed data.

The first error that I uncovered affected Dr. Bozzo's coefficient estimates for his pooled and random effects models. In these models, he neglected to apply the ρ transformation to the intercept terms in his models. Had he used the "textbook alternative," his coefficient estimate for the intercept term would have been off by a multiplicative constant, but otherwise his results would not have been affected. However, by using the procedure that allowed him to retain the first observation in each run, he created a situation in which his constant term was no longer constant. His failure to transform the intercept thus means that his results are incorrect. I have corrected this oversight in the results presented below.

Later in my analysis I uncovered a second error in his fixed effects model. His first two steps are carried out correctly, yielding an appropriate estimate of the

3. Response to UPS/USPS-T15-12 (March 22, 2000).

4. Dr. Bozzo uses the fixed effects estimator in TSP, which uses this procedure.

autocorrelation coefficient. He then applies the ρ transformation to his data. When he uses the TSP panel command to express the ρ-transformed data in terms of deviations from facility means and applies ordinary least squares to the doubly transformed data, he arrives at an incorrect result. The transformation that expresses data in terms of deviations from facility means in order to solve the fixed effects out of the model does not work on the transformed data. An alternative transformation contained in footnote 5 could have accomplished this.[5] Yet another alternative that would have avoided the error would have been to express the data in terms of deviations from facility means, and then apply the ρ transformation and use ordinary least squares to estimate the model coefficients.

Unfortunately, I uncovered the second error too late to allow me to rerun all of the models involved in Dr. Bozzo's original testimony and in my response to the Notice. All

---

5. In the standard procedure, the mean for some facility i is calculated as $\frac{\sum_{j=1}^{N_i} x_{jt}}{N_i}$, where N_i is the number of observations for facility i. This value is then subtracted from each of the x_{jt}'s. For Dr. Bozzo's transformed data, the correct value to subtract from each transformed variable x^*_{jt} is the quantity

$$\frac{\sum_{j=1}^{N_i} x^*_{jt}}{K_i\sqrt{1-\rho^2}+(N_i-K_i)(1-\rho)}$$, where K_i is the number of runs of data for facility i. The denominator in this expression equals the sum of the ρ transformed dummy variables.

of the results reported below involving facility specific fixed effects contain the same error.[6]

Correction of the first error had a substantial effect on the computation of the Hausman test comparing the fixed and random effects models. In most cases when the models are estimated correctly, the difference between the fixed and random effects covariance matrices turns out not to be positive definite, and hence it cannot be inverted.[7] This is something that is known to occur with the Hausman test when the asymptotic properties of the test fail. Hence, in most cases I am unable to use the Hausman test to determine whether the random effects hypothesis can be rejected.

4. Statistical Results

Table 1 presents the results of a series of hypothesis tests relating to Model A, for each of the MODS activities. These results are derived using Dr. Bozzo's procedures, and so they reflect both his failure to apply the autocorrelation adjustment to the constant terms in his models and his error in solving out the site specific fixed effects. The first two columns present specification test results for Dr. Bozzo's preferred specification without a correction for serial correlation of the error term. The second two columns present comparable results with correction for serial correlation. The latter results are preferred by Dr. Bozzo because of low values for the Durbin-Watson statistic, a diagnostic test for serial correlation. Within each set, the first column tests

6. In the models incorporating time period-specific fixed effects, I did not use the deviation from cell means transformation. Thus, these models do not reflect this particular error.

7. In his original testimony, Dr. Bozzo did not appear to have any problem computing the Hausman statistic. However, his ability to do so appears in most cases to have been an artifact of estimating the random effects model incorrectly.

the fixed effects model against the null hypothesis that the fixed effects are equal across sites. The null hypothesis in this case corresponds to Dr. Bozzo's "pooled" or OLS regression model. The relevant test in this case is an *F* test. The first item in each cell gives the calculated F-statistic. The second gives the p-value. The third gives the number of degrees of freedom for the F-Statistic. The second column in each set tests the fixed effects model against the null hypothesis that the site-specific effects are independently and identically distributed random variables. In this case, the relevant test is the Hausman test. The first entry in each cell gives the calculated chi-squared value. The second gives the p-value. The third gives the number of restrictions.

The results shown in Table 1 provide strong support for the fixed effects model. Regardless of whether or not a correction is made for serial correlation, the pooled model is strongly rejected in favor of the fixed effects model. The random effects model is similarly rejected in favor of the fixed effects model. In all cases, the alternative models are rejected by a large margin.

Table 2 presents comparable results for Model A after correcting for Dr. Bozzo's error in failing to apply the autocorrelation correction to the constant terms in his models. These revised results still reject the OLS model in favor of the fixed effects model, although the margins by which the OLS models are rejected are slightly reduced. Correction of the error has a marked effect, however, on the test of the fixed effects model against the random effects model when the two are estimated correcting for serial correlation. Correction of Dr. Bozzo's error results in a situation in which the Hausman statistic cannot be computed. It is for this reason that column 4 is blank. The same situation arose in all subsequent tests of the random effects specification, and for this reason I have omitted these tests from the results reported below.

Table 3 presents results for a series of hypothesis tests relating to Model B. The first column presents results for models without correction for serial correlation. The second presents results for models with correction for serial correlation. The latter results, and all comparable results reported below, reflect an appropriate autocorrelation adjustment of the constant term and a correction of Dr. Bozzo's first error. Because I estimated the version B models by explicitly including time period dummy variables rather than using the computational shortcut employed by Dr. Bozzo, these results are not subject to his second error.

These results test Model B with time-specific fixed effects against the alternative pooled regression model in which the time period-specific effects are equal across all time periods. They indicate that in a comparison between the pooled model and the fixed effects model, the fixed effects model is preferred.

Since the pooled, or OLS, regression model was rejected in favor of the fixed effects regression model for Models A and B, it is not possible, from the results presented thus far, to choose between Models A and B. For this reason, I have conducted an additional series of comparisons between these models and the more general Model C described above. Results of these comparisons are presented in Table 4. As in prior tables, the first set of columns present specification test results from regressions run without correction for serial correlation, while the second set presents results from regressions run with such correction. Within each set, the first column tests the fixed effects regression model for Model C against the null hypothesis of the pooled regression model. The second column tests Model C against the null hypothesis of Model A. The final column tests Model C against the null hypothesis of Model B. As noted above, the models involving facility specific fixed effects (Models A and C) are estimated using Dr. Bozzo's erroneous procedure.

The results presented in Table 4 support the fixed effects estimator for Model C over all of the alternatives for all MODS activities, except Manual Flats. The null hypothesis of the pooled regression model is rejected in favor of the fixed effects regression model for Model C in all cases. The null hypothesis of fixed effects for Model A is rejected in favor of fixed effects for Model C for all MODS activities, except Manual Flats. The null hypothesis of fixed effects for Model B is rejected in favor of fixed effects for Model C in all cases. Thus, from a strictly statistical standpoint, Model C with fixed effects estimation emerges as the clear winner for all but the Manual Flats MODS group. For Manual Flats, the Model A with fixed effects is the winner in the sense that this simpler specification cannot be rejected.

Table 5 compares the volume variabilities implied by these models. All of the variabilities are derived from models estimated with correction for serial correlation. Moving from the Model A fixed effects to Model B fixed effects raises the estimated volume variability in all cases except one. In some instances, the changes are fairly dramatic. The estimated volume variability for Manual Parcels, for example, goes from 0.522 to 0.641. That for Priority Mail goes from 0.522 to 0.641. The addition of time period-specific effects to Model A has the effect of reducing volume variability slightly in five of the nine MODS activities.

5. Interpretation and Discussion

The general conclusion, from a strictly statistical standpoint, is that the preferred model among those investigated is the fixed effects regression model for Model C, although the remaining error in the estimation of the fixed effects models leaves me unsure as to how valid or robust this finding really is. From this narrow viewpoint, there is little else to be said. The pattern of results presented raises some questions, however, about just what is going on in Dr. Bozzo's models.

In every instance in which a set of "dumb" variables is added to Dr. Bozzo's models, they appear to take statistically significant coefficients. In his original model, he included time trends and facility-specific fixed effects. In response to the Notice, I have added time period-specific fixed effects, and they also have turned out to be statistically significant. Even with time period-specific fixed effects and time trends, the models show evidence of serial correlation of the error term. One is left to wonder whether other as-yet unexplored possibilities might turn out to be statistically significant. Clusters of facility-time period interactions? Higher-order autocorrelation?

These changes in model specification sometimes have substantively important effects on estimates of volume variabilities. The question of what really belongs in the model thus appears to be an important one.

A clear implication of the tendency of these "dumb" variables to take statistically significant coefficients is that there is much going on in the labor hour data that is not explained well by the substantively important parts of Dr. Bozzo's model. This is hardly surprising, given the parsimoniousness of his specification. In addition to piece handlings, his model contains the manual ratio and his capital index – two variables that

I would regard as endogenous, rather than as independent determinants of labor demand. His wage variables are only weakly related to labor hours. Only the delivery points variable appears to play a strong role in the models.

In my direct testimony, I commented extensively on aspects of the Postal Service's response to volume changes that are nowhere represented in Dr. Bozzo's models.[8] Dr. Bozzo fails to account fully for the interactions among activities within a plant, and his analysis ignores the likelihood that the mix of sorting technologies within a plant will change systematically with growth in volume. From a theoretical standpoint, I thought it likely that his models were misspecified. The pattern of results presented above is consistent with that opinion. If his model is misspecified, it is likely that dummy variables, time trends, and serial correlation coefficients will pick up some of the effects of the omitted variables and, as a result, take statistically significant coefficients.

In this context, it is worth repeating the cautions expressed above regarding the unreliability of these statistical tests in the presence of misspecification. If Dr. Bozzo's models are misspecified, his coefficient estimates are biased and all of the tests reported above are unreliable. I believe that this is likely to be the case.

6. Are There Theoretical Reasons for Rejecting Model A?

The Notice invites discussion of the question "whether, even with the rejection of the hypotheses described in a), there may be theoretical grounds for concluding that a rejected model could provide a better estimate of variability than either model A or B." Notice at 3, ¶ f. Such grounds do exist. They have to do with the appearance on the right hand side of the regression equation of endogenous variables under the control of

8. UPS-T-1, pages 21-23, Tr. 27/12793-95.

the Postal Service. I alluded to them in my response to interrogatory USPS/UPS-T1-13(b), Tr. 27/12936-38. Portions of that response are worth repeating here:

> Many aspects of postal operations are likely to affect the structural relationship between mail processing labor costs and mail volume. However, many such aspects of postal operations -- including capital intensity, choice of sorting technology and the structure and organization of the mail processing network -- are under the control of the Postal Service, and likely themselves to change systematically in response to changes in mail volume. Simply including such explanatory variables in the regression model without accounting properly for their endogeneity is likely to lead to simultaneity bias. Moreover, even if the econometric problems associated with the inclusion of right hand side endogenous variables could be adequately resolved, the resulting structural model would produce incomplete results. It would capture the direct effects of volume on labor costs, holding other decision variables constant. However, it would exclude the indirect effects exerted by volume growth through its influence on these other decision variables.
>
> In such a situation the appropriate econometric model is a reduced form model that excludes from the right hand side all endogenous variables. The estimated coefficient on volume in such a model captures both the direct and indirect effects of volume on labor cost. The result is a more comprehensive measure of the volume variability of labor costs, and one that comes closer to meeting the requirements of the Commission.

The variability regressions presented by Dr. Bozzo contain a number of endogenous right hand side variables. These include the manual ratio, which measures the way in which the incoming mail stream is allocated between manual and automated sorting activities. They also include Dr. Bozzo's capital index, which clearly reflects Postal Service investment decisions. When Dr. Bozzo computes volume variabilities, he relies upon regression coefficients that control for the effects of changes in these endogenous variables and that effectively give the volume variability of labor hours *holding the manual ratio and the capital index constant.*

That said, the manual ratio and the capital index do not play a large role in Dr. Bozzo's analysis. Although they are generally significant in a statistical sense, their measured effects on labor hours are generally modest. Their modest role is probably in large part an artifact of the way in which they are measured. Dr. Bozzo's capital index is not limited to equipment relevant to a particular MODS activity, but rather represents a comprehensive measure of the amount of equipment present in the entire plant, with the plant thrown in as well. Given that so much equipment irrelevant to the particular MODS activity is included in this measure, it is somewhat surprising that there is a significant relationship at all.[9]

In my direct testimony on mail processing, I identified a number of ways in which the Postal Service responds to growth in volume that are not addressed by Dr. Bozzo's study. These include installation of automated processing activities in plants,[10] as well as expansions and/or modifications of plants, or the construction of new plants.[11] Variables describing these aspects of the Postal Service's response to volume changes do not appear explicitly in Dr. Bozzo's model. Since his analysis looks only at processing activities that are up and running, we never observe the installation and initiation of a new processing activity.

Dr. Bozzo's fixed effects coefficients measure aspects of labor hour demand that do not vary in response to quarter-to-quarter changes in piece handlings. There is disagreement, however, over whether they reflect, in whole or in part, Postal Service design and operational decisions that respond over a longer time period to expectations

9. OCA witness Smith has also criticized Dr. Bozzo's capital index for its reliance on accounting based depreciation rates that may have little or nothing to do with the actual loss of physical productivity that occurs over time. See OCA-T-4, page 34, line 16, through page 35, line 17, Tr. 27/13183-84.

10. UPS-T-1, pages 9-16, Tr. 27/12781-88.

11. UPS-T-1, pages 16-18, Tr. 27/12788-90.

regarding the volume of mail to be processed within a plant. OCA witness Smith noted the Commission's finding in Docket No. R97-1 that "the fixed effects in Dr. Bradley's study may represent effects that are both related and unrelated to volume."[12] Elsewhere in his testimony, Dr. Smith emphasized the importance of accounting appropriately for the characteristics of the longer-run expansion path mapping out the optimal combination of labor and capital for different levels of expected mail processing volume.[13] In my own direct testimony on mail processing costs, I discussed the way in which processing technology might change systematically in response to changes in mail volume.[14] Dr. Smith argues for use of Dr. Bozzo's "between" model on the argument that it is most likely to show the relationship between volume and cost as plant size varies.[15]

It is certainly reasonable to argue that when the Postal Service opens a new plant, it designs the plant to handle the volume of mail that plant is expected to process. It is also reasonable to expect anticipated volumes to trigger upgrading decisions, and to influence the characteristics of the plant that emerges from the upgrading process. It is likely, therefore, that when viewed in the cross-section, the different plants in the Postal Service's network represent different points on witness Smith's expansion path. In other words, they will be designed to accommodate different mail processing volumes. These design decisions are an important part of the Postal Service's long run

---

12. OCA-T-4, page 16, lines 1-2, Tr. 27/13165.

13. OCA-T-4, page 40, lines 14-18, Tr. 27/13189.

14. UPS-T-1, pages 11-14, Tr. 27/12783-86.

15. OCA-T-4, page 64, lines 7-12, Tr. 27/13213.

response to changes in volume. They will be reflected in plant size, layout, automation strategy, and many other attributes not explicitly represented in Dr. Bozzo's model.

Since volume-related plant design decisions change slowly and infrequently and are not represented explicitly in Dr. Bozzo's model, it is likely that they are captured in large part by his fixed effects. One can think conceptually of decomposing his fixed effects into two parts. One part would represent the truly fixed effects that would never change with volume. An example might be a location within an urban area. The remainder, however, would reflect volume-related aspects of plant design, such as the fact that in an urban area, the Postal Service will tend to build a large plant to process the large volumes of mail it can expect to have to process.

If it were true that volume-related design decisions account for most of the fixed effects estimated by Dr. Bozzo, these could be regarded as endogenous variables that are actually under the control of the Postal Service. In such a case, the argument presented above would apply. The appropriate measure of volume variability would reflect both the effects of long term volume growth on the number, size, and configuration of the plants in the processing network, as well as the effects of short term changes in the volume of mail processed within those plants. In such a situation, dropping the fixed effects could be regarded as the equivalent of running a reduced form model.

PRESIDING OFFICER'S INFORMATION REQUEST NO. 19

1. Summary

Presiding Officer's Information Request No. 19 ("the Request") cites my response to Interrogatory USPS/UPS-T1-10(b) (Tr. 27/12921-25) in which I discussed the incidence of errors in the MODS data used by Dr. Bozzo. In that response, I pointed out that the discussion of error rates contained in Dr. Bozzo's testimony understates the incidence of erroneous data, by failing to include in his count of errors observations lost because of missing or negative values for the variables key to his analysis. I also presented revised estimates of MODS data error rates that include the errors omitted from his calculations.

The Request notes that in addition to the types of errors described in my response to USPS/UPS-T1-10(b), there are also instances in the data in which Total Pieces Handled ("TPH") are greater than Total Pieces Fed ("TPF"). As explained by Dr. Bozzo, TPF represents the number of pieces of mail fed into a distribution operation, while TPH represents the number of pieces successfully sorted; the difference between the two, if any, consists of pieces jammed, pieces misfed, or pieces which for some other reason (such as the presence of unreadable addresses or barcodes) are incapable of being sorted.[16] By definition, TPF should always be greater than or equal to TPH. This, however, is not always the case in Dr. Bozzo's data set.

The Request asks a number of specific questions. It asks what meaning can be attached to non-positive values of TPH and TPF, and if there is any way to determine if positive values of TPH and TPF are infected by the sources of measurement error that give rise to the observed non-positive values. It asks also whether observations in which TPH is greater than TPF are indications of data errors. If so, it asks for an updated version of the table prepared in my response to USPS/UPS-T1-10(b) that reflects this additional source of error. It also asks whether there is any way to

16. USPS-T-15 at pages 50-52. Note that for manual operations, TPF and TPH are identical. They can differ only for automated activities.

determine whether observations in which TPF equals or exceeds TPH are infected by the same sources of measurement error that cause TPH to sometimes exceed TPF. It asks what the answers to the above questions imply for the variability analysis introduced by Dr. Bozzo, and whether the resulting bias is likely to be greater for the fixed effects model than for other models, such as the between model.

2. Patterns of Error within the MODS Data

In developing his econometric models of mail processing labor hour variability, Dr. Bozzo relies upon three variables drawn from the MODS data files: Labor Hours, Total Pieces Fed (TPF), and Total Pieces Handled (TPH). Logically, one would expect to see positive values for all three variables if a MODS activity were up and running at a site during a particular time period. Moreover, because of the definitions of TPH and TPF, one would expect that TPF should always be greater than or equal to TPH. Conversely, if a MODS activity is not present, values for all three variables should equal zero.

There are numerous instances in which the expected relationships among hours, TPH, and TPF do not hold. Hours are sometimes positive when TPH equals zero. The reverse relationship also holds. TPH and TPF frequently disagree in implausible ways. The source and significance of these errors is not clear. In his response to an interrogatory, Dr. Bozzo noted that manual parcel and priority volumes must be logged manually, and he suggested that gaps in the data for at least one specific site may have arisen because an in-plant support position was not filled.[17] In his direct testimony, he

17. Response to UPS/USPS-T15-13, Tr. 15/6387-88.

states that some sites appear to have systematically underreported TPF relative to TPH, although he gives no explanation of why this may have occurred.[18]

3. Interpretation of Non-Positive Values

In the data set produced by Dr. Bozzo, zero values have an ambiguous interpretation. They can represent either true zeros, or missing values. On the presumption that once activities are in place they tend to operate consistently rather than starting and stopping, I have treated runs of zeros at either the start or the end of the data for a site as true zeros, and runs of zeros that are embedded between positive values as missing values that represent failures of the MODS reporting system. How these missing values occur is not clear. The statements by Dr. Bozzo referred to above suggest that at times the reporting system simply breaks down. Apparently, these reporting failures can affect all of the variables used by Dr. Bozzo, or only some of them.

In principal, negative values have no proper place within the MODS data. However, they appear with some regularity. Their significance is not clear. I have seen instances in working with other data systems in which entries made to adjust prior period errors sometimes show up as negative values in the current period, and I suspect that some similar explanation may account, at least in part, for the presence of such negative values in the MODS data. The MODS manual does refer to procedures for making adjustments to prior period values.[19]

18. USPS-T-15, page 108, lines 4-6.

19. Management Operating Data System, Handbook M-32, Docket No. R97-1, USPS-LR-H-147, Section 432.1.

4. Are Other Observations Infected by the Problems Causing Non-Positive Values?

The problems that give rise to non-positive values for hours, TPH, or TPF could affect other apparently correct observations if the underlying reporting system records data at a finer level of aggregation than that used by Dr. Bozzo. His observations are quarterly and represent aggregations of Postal Service four-week accounting periods. If data were reported on a weekly basis, it would be possible for zero or negative values to appear in one of the four weeks of an accounting period and to be masked when data for the four weeks were aggregated together to produce accounting period totals.

It is clear that this possibility exists within Dr. Bozzo's data. He aggregated Postal Service accounting periods to arrive at his quarterly totals. I note also that the MODS manual appears to provide for the reporting of data at the day, tour, week, or accounting period level.[20] Unless there is some procedure within MODS that checks for errors before aggregating to a higher level, it is highly likely that some apparently correct observations contain hidden errors.

I know of no way from the presently available data to determine how extensive this problem is. The only way to determine the extent of this problem with any confidence would be to start with data at the finest level of aggregation available and check for errors at each stage of aggregation. Even such an extensive effort as this, however, would not necessarily identify the full extent of the problem. Reporting error and omissions could remain even within the finest level of aggregation maintained by the system. It is possible, for example, that at the end of each shift it is necessary to

20. Management Operating Data System, Handbook M-32, Docket No. R97-1, USPS-LR-H-147, Section 131.

enter piece counts from several different machines. For any number of different reasons, there could be a failure to report data for a particular machine that would be masked by the presence of data from the machines whose results were reported.

5. Do Observations in Which TPH Exceed TPF Represent Errors?

Observations in which TPH exceed TPF clearly represent errors. For automated operations, it is clear, even from Dr. Bozzo's testimony, that TPH should not exceed TPF. For manual operations, there is no meaningful distinction between TPH and TPF, and TPF should simply equal either TPH or zero.

In response to the request for an expanded version of "Table in Response to USPS/UPS-T-10(b)" (Request, page 2), I investigated the TPH and TPF data series in both the analysis sample used in the regressions and in the larger sample of observations provided by Dr. Bozzo. A summary of my findings is presented in Tables 6 and 7.

Table 6 reports the percent of sample observations that exhibit MODS data errors. Column 4 presents Dr. Bozzo's calculation of the fraction of observations that fail the threshold and productivity checks. Dr. Bozzo investigates errors only in the sample of observations used in his regression analysis. He ignores the fact that certain observations were omitted from the regression sample because of data errors. Column 5 expands the universe over which the threshold and productivity error rates are calculated to include in the "non-missing" set those observations that would have been non-missing but for bad MODS data. In response to the Request, Column 6 expands the types of errors which are investigated to include instances in which TPH > TPF.

Table 7 reports the prevalence of MODS data errors for each MODS group over the set of all observations that demonstrate the presence of the MODS activity. Detecting the presence of MODS activity is complicated by the fact that Dr. Bozzo's data codes both missing values and non-present (truly zero) activities as zero. The MODS activity is considered to be present if at least one of the three MODS variables (TPH, TPF, or Hours) is strictly positive, or if at least one of the three MODS variables is an intermittent non-positive number, as explained in the workpapers accompanying my original testimony.

6. Are Other Observations Infected by the Problems that Cause TPH to Exceed TPF?

As explained above for non-positive values, the problems that cause TPH to exceed TPF could affect other apparently correct observations. If the underlying reporting system records data at a finer level of aggregation than that used by Dr. Bozzo, as described above, then it may well be that data errors are masked when data are aggregated to produce accounting period totals.

7. Are These Data Errors Likely to Produce Greater or Lesser Bias in the Fixed Effects Model Than in Other Models?

Measurement error in the right hand side variables of the regression model destroys the statistical properties of the panel estimators. While there exists some simulation evidence to suggest that there may be a trade-off in the relative bias of the different panel estimators, there is in general no way to determine which model is likely to produce greater or lesser bias.[21]

21. See E. Biorn, "The Bias of Some Estimators for Panel Data Models with Measurement Errors," Empirical Economics, vol. 17, 1992, pp. 51-66.

Table 1

Specification Tests for Model A

F-statistic comparing OLS to FE, Hausman test statistic comparing FE to RE.

	Without Correction for Serial Correlation			With Correction for Serial Correlation		
MODS Group	**OLS vs FE [1]**		**FE vs RE [2]**	**OLS vs FE [3]**		**FE vs RE [4]**
OCR	35.044		136.807	7.420		110.223
	(0.000)		(0.000)	(0.000)		(0.000)
	288	4762	38	288	4761	38
LSM	18.987		90.161	6.569		76.316
	(0.000)		(0.000)	(0.000)		(0.000)
	272	3583	38	272	3583	38
BCS	35.455		69.768	9.145		72.104
	(0.000)		(0.001)	(0.000)		(0.001)
	296	5056	38	296	5055	38
Manual Letters	44.211		191.995	10.631		168.657
	(0.000)		(0.000)	(0.000)		(0.000)
	298	5163	38	298	5162	38
FSM	45.575		172.756	11.660		76.862
	(0.000)		(0.000)	(0.000)		(0.000)
	234	4084	38	234	4084	38
Manual Flats	39.858		258.642	9.145		123.051
	(0.000)		(0.000)	(0.000)		(0.000)
	276	4564	38	276	4564	38
SPBS	53.546		60.420	15.917		50.547
	(0.000)		(0.001)	(0.000)		(0.015)
	93	1445	31	93	1445	31
Manual Parcels	41.583		119.299	12.898		83.131
	(0.000)		(0.000)	(0.000)		(0.000)
	180	2812	31	180	2811	31
Priority	27.197		108.282	9.642		83.057
	(0.000)		(0.000)	(0.000)		(0.000)
	199	3010	31	199	3009	31

Notes:

1. Model A is specified as $y_{it} = \gamma_t + \alpha_i + X_{it}\beta + \varepsilon_{it}$, where $\gamma_t = \gamma$ for all t.
2. Columns [1] and [3]: F-statistics shown on first line of each cell. P-values shown in parentheses on second line. Number of restrictions and degrees of freedom shown on third line.
3. Columns [2] and [4]: Hausman test statistics shown on first line of each cell. P-values shown in parentheses on second line. Degrees of freedom shown on third line.

Table 2
Specification Tests for Model A
With Correction to Dr. Bozzo's FGLS Transformation
F-statistic comparing OLS to FE, Hausman test statistic comparing FE to RE.

MODS Group	Without Correction for Serial Correlation: OLS vs FE [1]	Without Correction for Serial Correlation: FE vs RE [2]	With Correction for Serial Correlation: OLS vs FE [3]	With Correction for Serial Correlation: FE vs RE [4]
OCR	35.044 (0.000) 288 | 4762	136.807 (0.000) 38	7.354 (0.000) 288 | 4761	
LSM	18.987 (0.000) 272 | 3583	90.161 (0.000) 38	6.463 (0.000) 272 | 3583	
BCS	35.455 (0.000) 296 | 5056	69.768 (0.001) 38	9.029 (0.000) 296 | 5055	
Manual Letters	44.211 (0.000) 298 | 5163	191.995 (0.000) 38	10.561 (0.000) 298 | 5162	
FSM	45.575 (0.000) 234 | 4084	172.756 (0.000) 38	11.547 (0.000) 234 | 4084	
Manual Flats	39.858 (0.000) 276 | 4564	258.642 (0.000) 38	9.134 (0.000) 276 | 4564	
SPBS	53.546 (0.000) 93 | 1445	60.420 (0.001) 31	15.841 (0.000) 93 | 1445	
Manual Parcels	41.583 (0.000) 180 | 2812	119.299 (0.000) 31	12.865 (0.000) 180 | 2811	
Priority	27.197 (0.000) 199 | 3010	108.282 (0.000) 31	9.655 (0.000) 199 | 3009	

Notes:

1. Model A is specified as $y_{it} = \gamma_\tau + \alpha_i + X_{it}\beta + \varepsilon_{it}$, where $\gamma_t = \gamma$ for all t.
2. Columns [1] and [3]: F-statistics shown on first line of each cell. P-values shown in parentheses on second line. Number of restrictions and degrees of freedom shown on third line.
3. Column [2]: Hausman test statistics shown on first line of each cell. P-values shown in parentheses on second line. Degrees of freedom shown on third line.
4. Column [4]: It was not possible to compute the Hausman statistic in these instances.

Table 3

Specification Tests for Model B

With Correction to Dr. Bozzo's FGLS Transformation

F-statistic comparing OLS to FE, Hausman test statistic comparing FE to RE.

MODS Group	Without Correction for Serial Correlation		With Correction for Serial Correlation	
	OLS vs FE		OLS vs FE	
OCR	5.973		4.963	
	(0.000)		(0.000)	
	18	5037	18	5036
LSM	2.242		4.451	
	(0.002)		(0.000)	
	18	3842	18	3842
BCS	3.188		7.925	
	(0.000)		(0.000)	
	18	5339	18	5338
Manual Letters	4.323		22.734	
	(0.000)		(0.000)	
	18	5448	18	5447
FSM	5.847		15.301	
	(0.000)		(0.000)	
	18	4305	18	4305
Manual Flats	2.489		4.047	
	(0.000)		(0.000)	
	18	4827	18	4827
SPBS	2.294		8.293	
	(0.002)		(0.000)	
	18	1525	18	1525
Manual Parcels	2.177		6.527	
	(0.003)		(0.000)	
	18	2979	18	2978
Priority	2.895		9.891	
	(0.000)		(0.000)	
	18	3196	18	3195

Notes:

1. Model B is specified as $y_{it} = \alpha_i + \gamma_t + X_{it}\beta + \varepsilon_{it}$, where $\alpha_i = \alpha$ for all i.

2. F-statistics shown on first line of each cell. P-values shown in parentheses on second line. Number of restrictions and degrees of freedom shown on third line.

Table 4

Specification Tests Comparing Models With and Without Time-Specific Effects and Site-Specific Effects

MODS Group	Without Correction for Serial Correlation						With Correction for Serial Correlation					
	OLS *vs* Model C FE		Model A FE *vs* Model C FE		Model B FE *vs* Model C FE		OLS *vs* Model C FE		Model A FE *vs* Model C FE		Model B FE *vs* Model C FE	
	[1]		[2]		[3]		[4]		[5]		[6]	
OCR	33.825		2.863		35.246		7.190		1.877		7.367	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.028)		(0.000)	
	301	4749	13	4749	288	4749	301	4748	13	4748	288	4748
LSM	18.415		3.231		19.051		6.566		4.023		6.544	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
	285	3570	13	3570	272	3570	285	3570	13	3570	272	3570
BCS	34.697		6.341		35.938		9.069		6.179		9.073	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
	309	5043	13	5043	296	5043	309	5042	13	5042	296	5042
Manual Letters	43.338		7.284		45.009		10.248		6.494		10.192	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
	311	5150	13	5150	298	5150	311	5149	13	5149	298	5149
FSM	43.471		2.273		45.737		11.241		4.903		11.473	
	(0.000)		(0.006)		(0.000)		(0.000)		(0.000)		(0.000)	
	247	4071	13	4071	234	4071	247	4071	13	4071	234	4071
Manual Flats	38.160		1.327		39.893		8.801		1.290		9.163	
	(0.000)		(0.189)		(0.000)		(0.000)		(0.211)		(0.000)	
	289	4551	13	4551	276	4551	289	4551	13	4551	276	4551
SPBS	48.365		3.318		54.735		14.702		3.045		16.285	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
	106	1432	13	1432	93	1432	106	1432	13	1432	93	1432
Manual Parcels	39.497		3.625		42.004		12.385		4.514		12.955	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
	193	2799	13	2799	180	2799	193	2798	13	2798	180	2798
Priority	26.242		4.802		27.685		9.604		6.329		9.945	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
	212	2997	13	2997	199	2997	212	2996	13	2996	199	2996

Notes:

1. The general model, denoted as Model C, is specified as $y_{it} = \alpha_i + \gamma_t + X_{it}\beta + \varepsilon_{it}$, where α_i is a site-specific effect and γ_t is a time-specific effect.

   Model A, or Bozzo's model, is specified as $y_{it} = \gamma_\tau + \alpha_i + X_{it}\beta + \varepsilon_{it}$, where $\gamma_t = \gamma$ for all t.

   Model B is specified as $y_{it} = \alpha_\iota + \gamma_t + X_{it}\beta + \varepsilon_{it}$, where $\alpha_i = \alpha$ for all i.

   The OLS model is specified as $y_{it} = (\alpha + \gamma) + X_{it}\beta + \varepsilon_{it}$.

2. F-statistics shown on first line of each cell. P-values shown in parentheses on second line. Number of restrictions and degrees of freedom shown on third line.

Table 5

Estimated Volume Variabilities

MODS Group	Bozzo's Results Model A Fixed Effects FGLS	Corrected FGLS Transformation Model B Fixed Effects FGLS	Model C Fixed Effects FGLS
	[1]	[2]	[3]
OCR	0.751 (0.038)	0.847 (0.038)	0.735 (0.039)
LSM	0.955 (0.021)	0.932 (0.026)	0.970 (0.022)
BCS	0.895 (0.030)	0.919 (0.028)	0.867 (0.030)
FSM	0.817 (0.026)	0.926 (0.022)	0.837 (0.026)
Manual Flats	0.772 (0.027)	0.833 (0.025)	0.766 (0.028)
Manual Letters	0.735 (0.024)	0.825 (0.024)	0.733 (0.024)
SPBS	0.641 (0.045)	0.742 (0.043)	0.654 (0.046)
Manual Parcels	0.522 (0.028)	0.641 (0.032)	0.513 (0.028)
Priority	0.522 (0.025)	0.641 (0.026)	0.507 (0.025)

Notes:

1. Random effects estimation for site-specific error component.
2. Standard errors shown in parentheses.

Table 6
Expanded Table in Response to USPS-UPS-T1-10

MODS Group	Non-Missing	Threshold	Threshold and Productivity	% of Observations Exhibiting Data Errors		
				Ignoring Non-Positive MODS Data	Accounting for Non-Positive MODS data	Accounting for TPH > TPF
	(1)	(2)	(3)	(4)	(5)	(6)
OCR	6642	6637	6493	2.24%	3.19%	3.40%
LSM	5155	5149	5126	0.56%	6.94%	7.59%
BCS	6882	6880	6777	1.53%	1.54%	2.98%
FSM	5441	5441	5423	0.33%	1.00%	9.46%
Manual Flats	6910	6910	6416	7.15%	7.16%	
Manual Letters	6910	6910	6820	1.30%	1.32%	
SPBS	2241	2236	2210	1.38%	8.45%	10.85%
Manual Parcels	5831	5621	4709	19.24%	28.07%	
Priority	5713	5640	4992	12.62%	22.04%	

Notes and Sources:

1. Data from USPS-T-15 (revised 3/22/00) and Reg9398.xls in USPS-LR-I-107.
2. "% of Observations Exhibiting Data Errors" columns show the percentage of observations exhibiting gross data errors when properly accounting for true missing value and bad TPH or work hours data.
3. Column (5) counts as bad data observations with complete non-MODS data, but non-positive values for either TPH or HRS.
4. Column (6) counts as bad usable observations (after the threshold and productivity scrubs) with TPH > TPF.

Table 7
MODS Data Quality

Description	OCR	LSM	BCS	Manual Letters	FSM	Manual Flats	SPBS	Manual Parcels	Priority
Sample Size	7140	6132	7472	7570	5963	7556	2771	7274	6908
TPH > 0, HRS ≤ 0	0.08	0.08	0.03	0.03	0.07	0.01	0.29	3.18	0.84
TPH ≤ 0, HRS > 0	0.77	6.21	0.24	0.21	0.62	0.21	7.33	7.23	9.87
TPH ≤ 0, HRS ≤ 0	0.53	3.02	0.16	0.23	1.58	0.09	5.52	2.63	1.84
TPH > 0, HRS > 0 Threshold failure	0.08	0.13	0.15	0.03	0.10	0.04	0.18	2.98	1.09
TPH > 0, HRS > 0 Productivity failure[1]	2.10	0.42	1.55	1.59	0.59	7.28	1.16	16.00	10.54
TPH > TPF	0.41	0.83	1.57		8.47		2.17		
TPF > 0, TPH =/ TPF				21.10		18.69		3.46	4.91
Overall % of MODS Data Exhibiting Error	3.95	10.62	3.57	22.84	11.34	24.38	16.46	32.05	27.26

Notes:

1. Productivity defined using original MODS data. Productivity bounds taken from USPS-T-15.
2. Threshold failure defined as hours greater than zero, but less than 40.